

# How to use MSstats as an external tool in Skyline

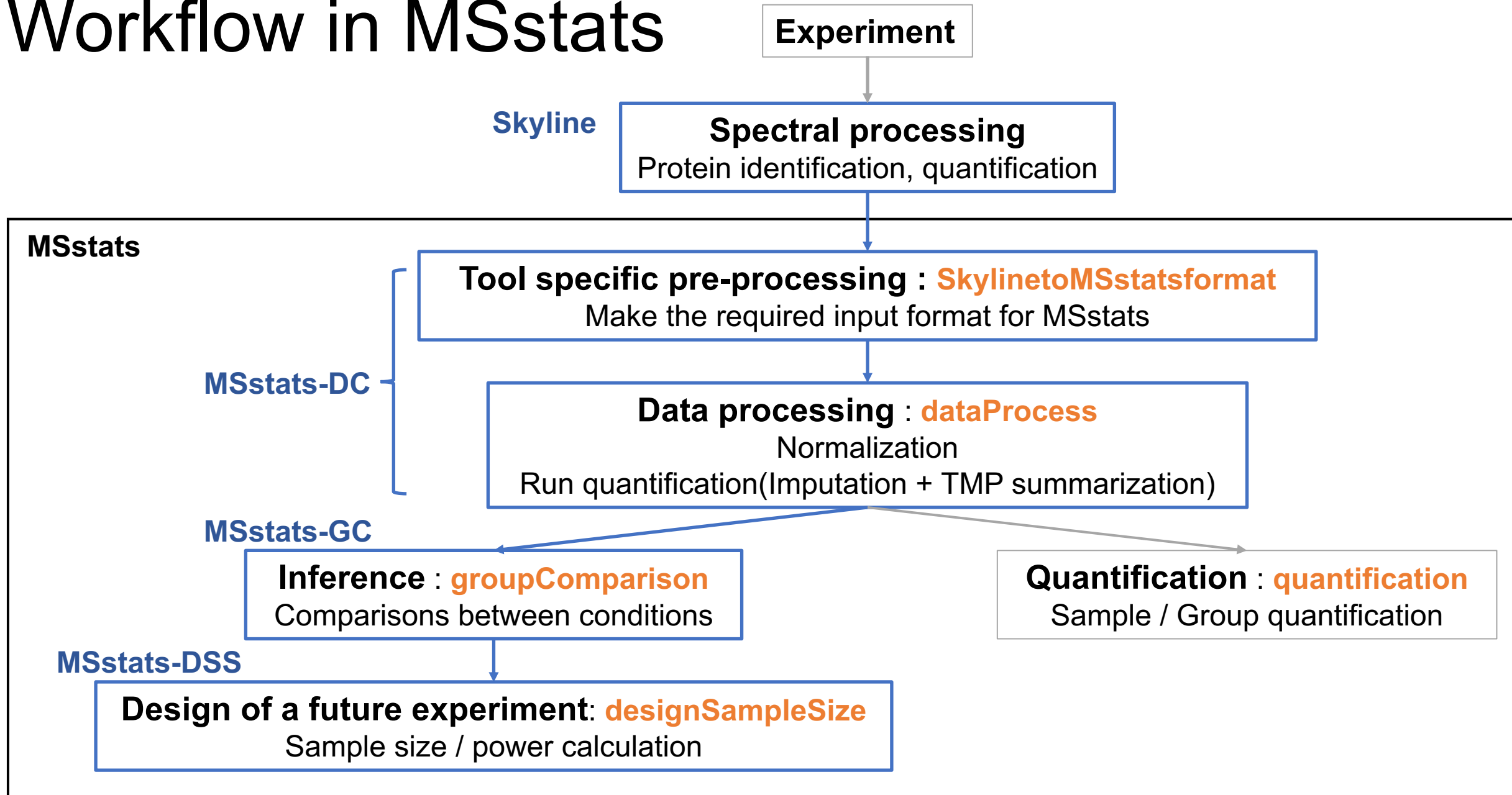
**Meena Choi**

Two-day short course : Case studies in Quantitative Proteomics  
at ASMS 2018

# Take-home message

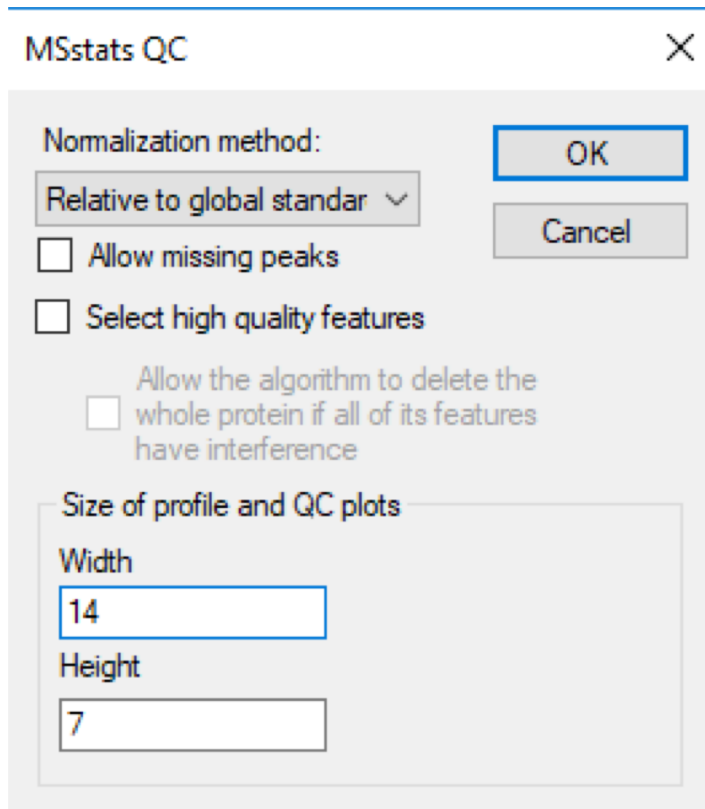
1. How to run MSstats external tool in Skyline
2. How to correctly annotate your experiments
3. Different parameter settings have different result
4. Other workflow with MSstats R package

# Workflow in MSstats



# 1. How to run MSstats external tool in Skyline

Heart failure with rat,  
Label-free SRM  
'Group studies' in tutorial



MSstats QC

Normalization method:  
Relative to global standard

☐ Allow missing peaks

☐ Select high quality features

☐ Allow the algorithm to delete the whole protein if all of its features have interference

Size of profile and QC plots

Width  
14

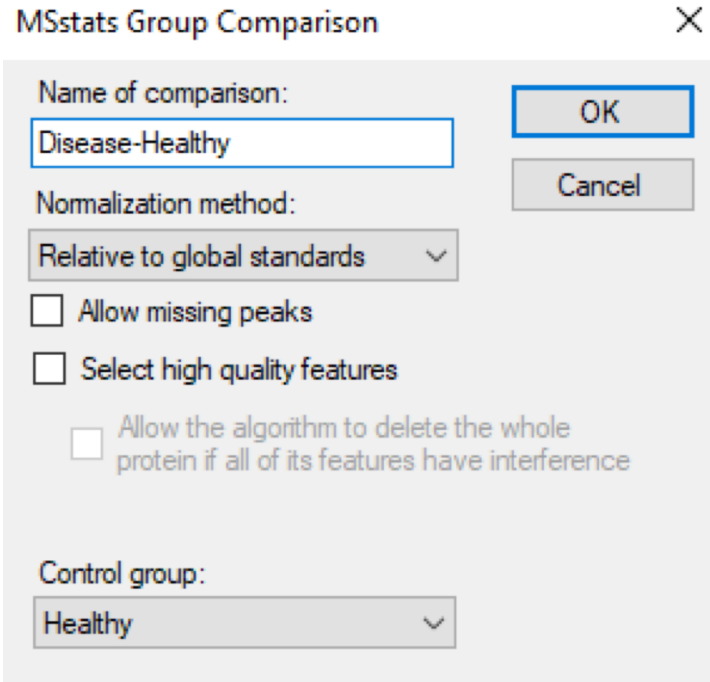
Height  
7

OK  
Cancel

- Tool > MSstats > data processing
- Normalization : 'Relative to global standard' with one peptide, VVD
- Data is processed for statistical analysis and three pdfs with visualization are generated in the folder with skyline file.
  - Profile plot
  - QC plot
  - Condition plot
- Video

# 1. How to run MSstats external tool in Skyline

Heart failure with rat,  
Label-free SRM  
'Group studies' in tutorial



MSstats Group Comparison

Name of comparison:  
Disease-Healthy

Normalization method:  
Relative to global standards

☐ Allow missing peaks

☐ Select high quality features

☐ Allow the algorithm to delete the whole protein if all of its features have interference

Control group:  
Healthy

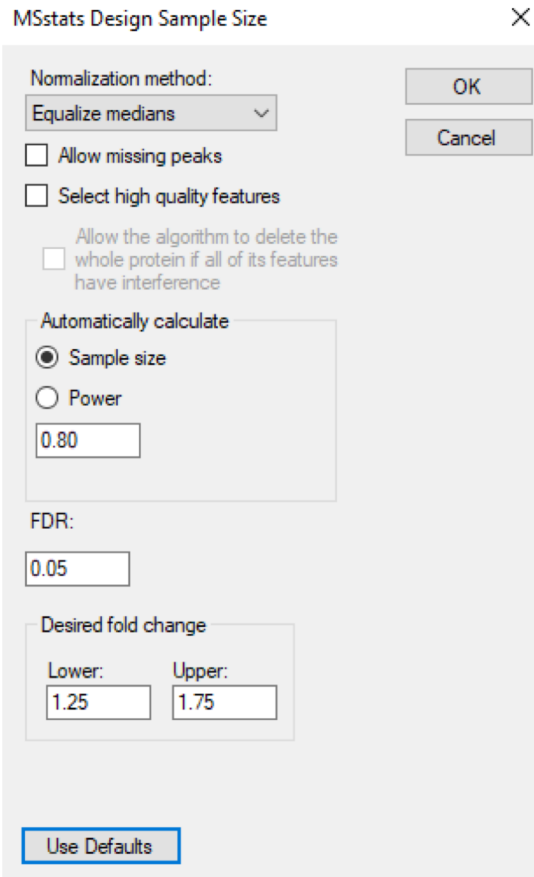
OK

Cancel

- Tool > MSstats > group comparison
- Normalization : 'Relative to global standard' with one peptide, VVD
- Set up which comparison you want.
- csv file with testing result and pdfs with visualization for results are generated in the folder with skyline file.
  - Volcano plot
  - Condition plot
- Video

# 1. How to run MSstats external tool in Skyline

Heart failure with rat,  
Label-free SRM  
'Group studies' in tutorial




The screenshot shows the 'MSstats Design Sample Size' dialog box. It has a title bar with a close button (X). The dialog contains several settings: 'Normalization method:' is set to 'Equalize medians' with a dropdown arrow; there are checkboxes for 'Allow missing peaks' (unchecked), 'Select high quality features' (unchecked), and 'Allow the algorithm to delete the whole protein if all of its features have interference' (unchecked); under 'Automatically calculate', 'Sample size' is selected with a radio button and 'Power' is unselected; there is a text input field for 'Sample size' containing '0.80'; 'FDR:' is set to '0.05' in a text input field; 'Desired fold change' has 'Lower:' set to '1.25' and 'Upper:' set to '1.75' in text input fields; and a 'Use Defaults' button at the bottom left. 'OK' and 'Cancel' buttons are at the top right.

- Tool > MSstats > design sample size
- Normalization : 'Relative to global standard' with one peptide, VVD
- Change values as you want.
- csv file with sample size calculation or power and pdf for visualization is generated in the folder with skyline file.
- Video

# Statistical analysis for iPRG data with MSstats external tool in Skyline

Secure | <https://panoramaweb.org/project/MacCoss/brendan/manuscripts/iPRG%202015/begin.view>

 iPRG 2015 Re-Processed Reports.zip (113.2 MB) MSstats reports : can be used in R

### Targeted MS Runs

1 - 9 of 9

Flag	File	Created	Proteins	Peptides	Small Molecules	Precursors	Transitions	Download
	<a href="#">iPRG_10ppm_2rt_05cut_nodup_2016-08-01_05-46-24.sky.zip</a>	2016-08-01	5,017	33,805	0	38,224	114,672	<a href="#">Download (3 GB)</a>
	<a href="#">iPRG_10ppm_2rt_05cut_nosingle_2016-08-01_11-45-26.sky.zip</a>	2016-08-01	3,320	32,108	0	36,494	109,482	<a href="#">Download (3 GB)</a>
	<a href="#">iPRG_10ppm_2rt_15cut_nodup_2016-08-05_05-22-38.sky.zip</a>	2016-08-05	4,451	32,006	0	36,320	108,960	<a href="#">Download (3 GB)</a>
	<a href="#">iPRG_10ppm_2rt_15cut_nosingle_2016-08-01_13-54-08.sky.zip</a>	2016-08-01	3,097	30,652	0	34,937	104,811	<a href="#">Download (3 GB)</a>
	<a href="#">iPRG_10ppm_2rt_50cut_nodup_2016-08-02_00-09-12.sky.zip</a>	2016-08-02	3,901	30,155	0	34,328	102,984	<a href="#">Download (3 GB)</a>
	<a href="#">iPRG_10ppm_2rt_50cut_nosingle_2016-08-02_23-25-24.sky.zip</a>	2016-08-03	2,935	29,189	0	33,337	100,011	<a href="#">Download (3 GB)</a>
	<a href="#">iPRG_10ppm_2rt_95cut_nodup_2016-08-16_16-37-53.sky.zip</a>	2016-08-16	3,547	28,295	0	32,218	96,654	<a href="#">Download (3 GB)</a>
	<a href="#">iPRG_10ppm_2rt_95cut_nosingle_2016-08-03_04-44-06.sky.zip</a>	2016-08-03	2,828	27,576	0	31,479	94,437	<a href="#">Download (3 GB)</a>
	<a href="#">iPRG_10ppm_2rt_template_2016-08-01_05-37-39.sky.zip</a>	2016-08-01	0	0	0	0	0	<a href="#">Download (2 KB)</a>

Use today

### Targeted MS Experiment

**ABRF Proteome Informatics Research Group (iPRG) 2015 Study: Detection of differentially abundant proteins in label-free quantitative LC-MS/MS experiments**  
[Edit] [Delete] [More Details...]

Panorama Public link: <https://panoramaweb.org/iPRG-2015.url> [Share](#)

**Organism:** Yeast  
**Instrument:** Thermo Scientific Q Exactive  
**Spikeln:** Yes

# Statistical analysis for iPRG data with MSstats external tool in Skyline

MSstats QC

Normalization method:  
Equalize medians   
☐ Allow missing peaks  
☐ Select high quality features  
☐ Allow the algorithm to delete the whole protein if all of its features have interference  
Size of profile and QC plots  
Width  
7  
Height  
7



MSstats Group Comparison

Name of comparison:  
Sample2-Sample1   
Normalization method:  
Equalize medians   
☐ Allow missing peaks  
☐ Select high quality features  
☐ Allow the algorithm to delete the whole protein if all of its features have interference  
Control group:  
sample1  
Select group(s) to compare against:  
sample2  
sample3  
sample4



MSstats Design Sample Size



Normalization method:  
Equalize medians   
☐ Allow missing peaks  
☐ Select high quality features  
☐ Allow the algorithm to delete the whole protein if all of its features have interference  
Automatically calculate  
☒ Sample size  
☐ Power  
0.80  
FDR:  
0.05  
Desired fold change  
Lower: 1.25 Upper: 1.75



## 2. How to correctly annotate your experiments

- Time- course
- Paired-design
- Case-control

# Time-course

## Annotation

Individual 1  
(BioReplicate1)

**Time 1**

**Time 2**

**Time 3**

Condition : Time1  
BioReplicate : 1  
Run : 1

Condition : Time2  
BioReplicate : 1  
Run : 6

Condition : Time3  
BioReplicate : 1  
Run : 8

Individual 2  
(BioReplicate2)

Condition : Time1  
BioReplicate : 2  
Run : 4

Condition : Time2  
BioReplicate : 2  
Run : 9

Condition : Time3  
BioReplicate : 2  
Run : 2

Individual 3  
(BioReplicate3)

Condition : Time1  
BioReplicate : 3  
Run : 3

Condition : Time2  
BioReplicate : 3  
Run : 5

Condition : Time3  
BioReplicate : 3  
Run : 7

Condition	BioReplicate	Run
Time1	1	1
Time2	1	6
Time3	1	8
Time1	2	4
Time2	2	9
Time3	2	2
Time1	3	3
Time2	3	5
Time3	3	7

\*The injection order (MS run) is randomized.

# Paired-design

	Healthy tissue1	Healthy tissue2	Lung cancer
Individual 1 (BioReplicate1)	Condition : Healthy1 BioReplicate : 1 Run : 1	Condition : Healthy2 BioReplicate : 1 Run : 6	Condition : cancer BioReplicate : 1 Run : 8
Individual 2 (BioReplicate2)	Condition : Healthy1 BioReplicate : 2 Run : 4	Condition : Healthy2 BioReplicate : 2 Run : 9	Condition : cancer BioReplicate : 2 Run : 2
Individual 3 (BioReplicate3)	Condition : Healthy1 BioReplicate : 3 Run : 3	Condition : Healthy2 BioReplicate : 3 Run : 5	Condition : cancer BioReplicate : 3 Run : 7

## Annotation

Condition	BioReplicate	Run
Healthy tissue1	1	1
Healthy tissue2	1	6
Lung cancer	1	8
Healthy tissue1	2	4
Healthy tissue2	2	9
Lung cancer	2	2
Healthy tissue1	3	3
Healthy tissue2	3	5
Lung cancer	3	7

\*The injection order (MS run) is randomized.

# Case-control

Healthy group

Lung cancer

Individual 1  
(BioReplicate1)

Condition : Healthy  
BioReplicate : 1  
Run : 1

Individual 2  
(BioReplicate2)

Condition : Healthy  
BioReplicate : 2  
Run : 4

Individual 3  
(BioReplicate3)

Condition : Healthy  
BioReplicate : 3  
Run : 3

Individual 4  
(BioReplicate4)

Condition : cancer  
BioReplicate : 4  
Run : 5

Individual 5  
(BioReplicate5)

Condition : cancer  
BioReplicate : 5  
Run : 6

Individual 6  
(BioReplicate6)

Condition : cancer  
BioReplicate : 6  
Run : 2

Annotation

Condition	BioReplicate	Run
Healthy	1	1
Healthy	2	4
Healthy	3	3
Lung cancer	4	5
Lung cancer	5	6
Lung cancer	6	2

\*The injection order (MS run) is randomized.

# Case-control with tech replicates

	Healthy group	Lung cancer
Individual 1 (BioReplicate1)	Condition : Healthy BioReplicate : 1 Run : 1, 8, 15	
Individual 2 (BioReplicate2)	Condition : Healthy BioReplicate : 2 Run : 4, 11, 14	
Individual 3 (BioReplicate3)	Condition : Healthy BioReplicate : 3 Run : 3, 10, 17	
Individual 4 (BioReplicate4)		Condition : cancer BioReplicate : 4 Run : 5, 9, 13
Individual 5 (BioReplicate5)		Condition : cancer BioReplicate : 5 Run : 6, 12, 16
Individual 6 (BioReplicate6)		Condition : cancer BioReplicate : 6 Run : 2, 7, 18

## Annotation

Condition	BioReplicate	Run
Healthy	1	1
Healthy	1	8
Healthy	1	15
Healthy	2	4
Healthy	2	11
Healthy	2	14
Healthy	3	3
Healthy	3	10
Healthy	3	17
Lung cancer	4	5
Lung cancer	4	9
Lung cancer	4	13
Lung cancer	5	6
Lung cancer	5	12
Lung cancer	5	16
Lung cancer	6	2
Lung cancer	6	7
Lung cancer	6	18

\*The injection order (MS run) is randomized.

# Example with different annotation

## iPRG2015

Case1 – correct for this experiment

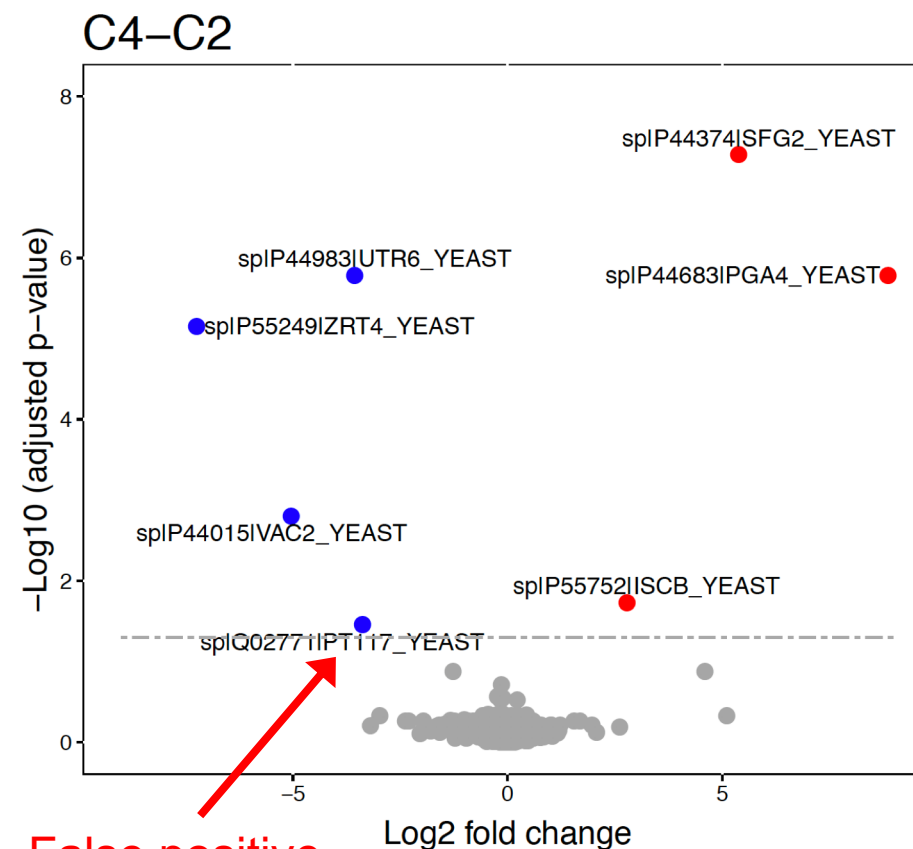
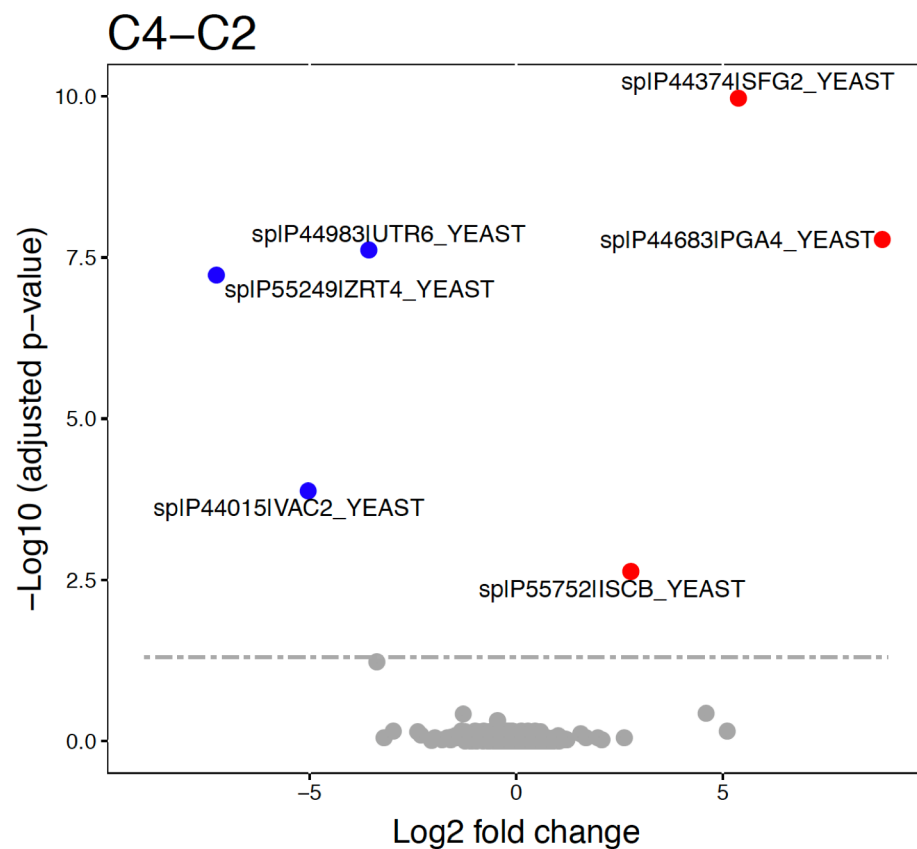
Run	Condition	BioReplicate
JD_06232014_sample1-A.raw	Condition1	1
JD_06232014_sample1_B.raw	Condition1	1
JD_06232014_sample1_C.raw	Condition1	1
JD_06232014_sample2_A.raw	Condition2	2
JD_06232014_sample2_B.raw	Condition2	2
JD_06232014_sample2_C.raw	Condition2	2
JD_06232014_sample3_A.raw	Condition3	3
JD_06232014_sample3_B.raw	Condition3	3
JD_06232014_sample3_C.raw	Condition3	3
JD_06232014_sample4_B.raw	Condition4	4
JD_06232014_sample4-A.raw	Condition4	4
JD_06232014_sample4_C.raw	Condition4	4

Case2 – incorrect for this experiment,  
common mistake

Run	Condition	BioReplicate
JD_06232014_sample1-A.raw	Condition1	1
JD_06232014_sample1_B.raw	Condition1	2
JD_06232014_sample1_C.raw	Condition1	3
JD_06232014_sample2_A.raw	Condition2	1
JD_06232014_sample2_B.raw	Condition2	2
JD_06232014_sample2_C.raw	Condition2	3
JD_06232014_sample3_A.raw	Condition3	1
JD_06232014_sample3_B.raw	Condition3	2
JD_06232014_sample3_C.raw	Condition3	3
JD_06232014_sample4_B.raw	Condition4	2
JD_06232014_sample4-A.raw	Condition4	1
JD_06232014_sample4_C.raw	Condition4	3

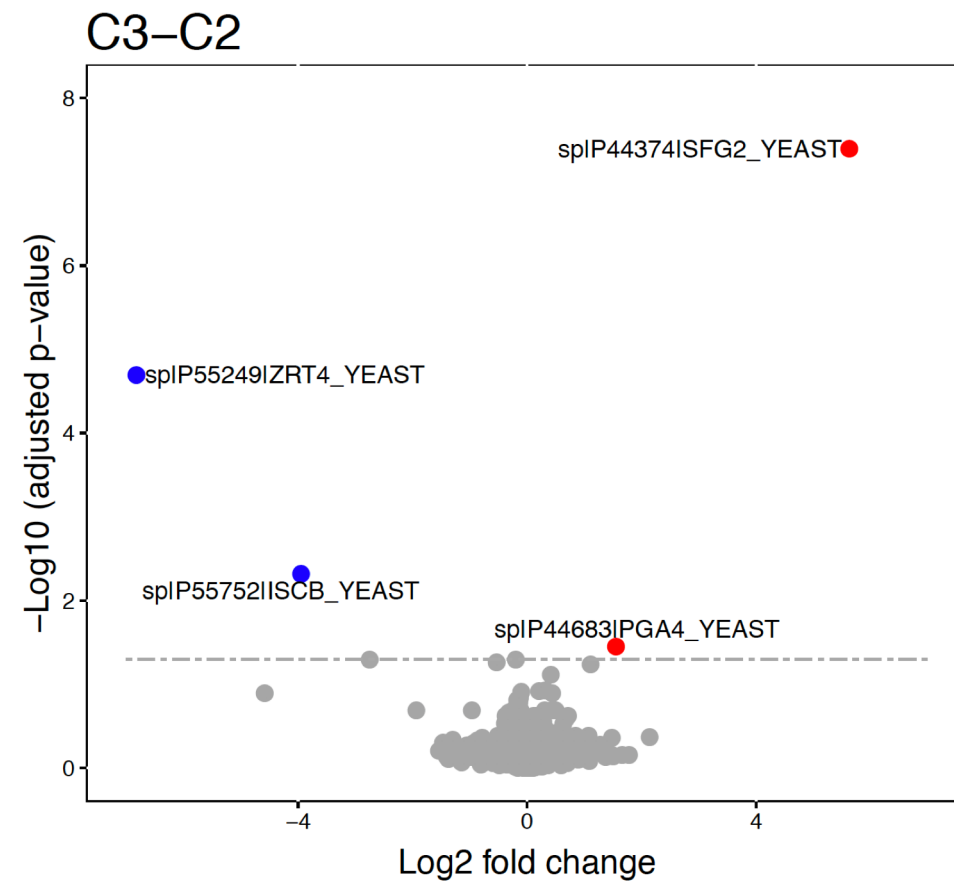
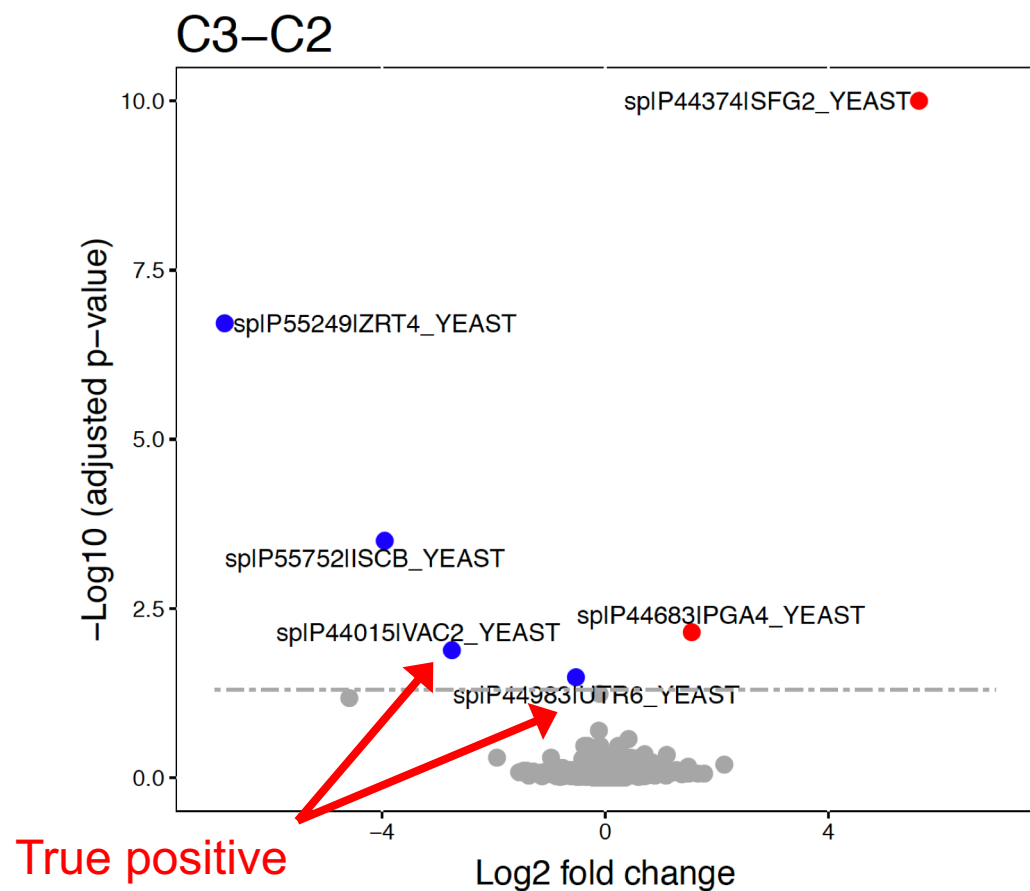
# Example with different annotation

N of significant proteins	Case1-correct	Case2-incorrect
Background protein	1	4
Spike protein	28 out of 30	26 out of 30



# Example with different annotation

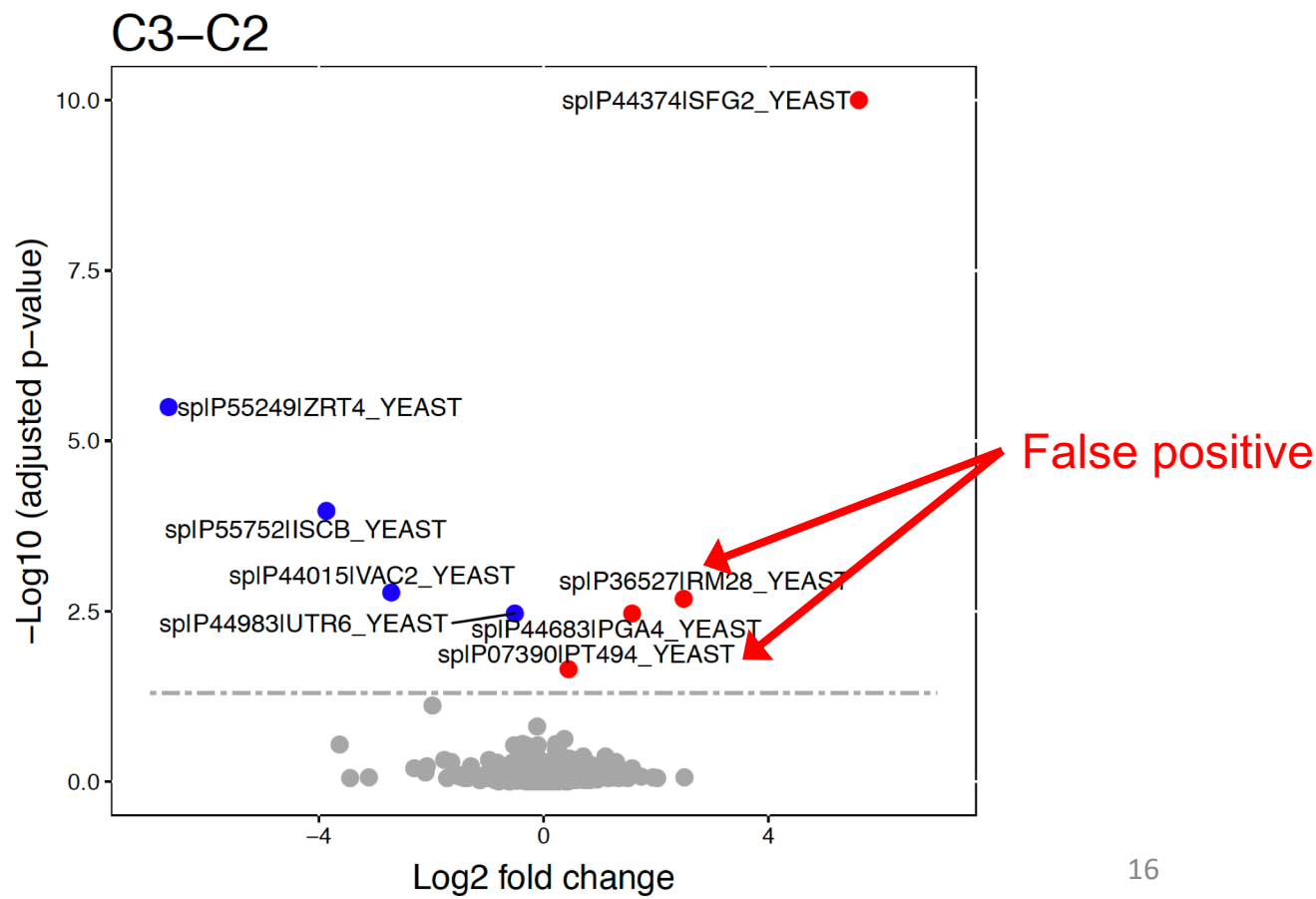
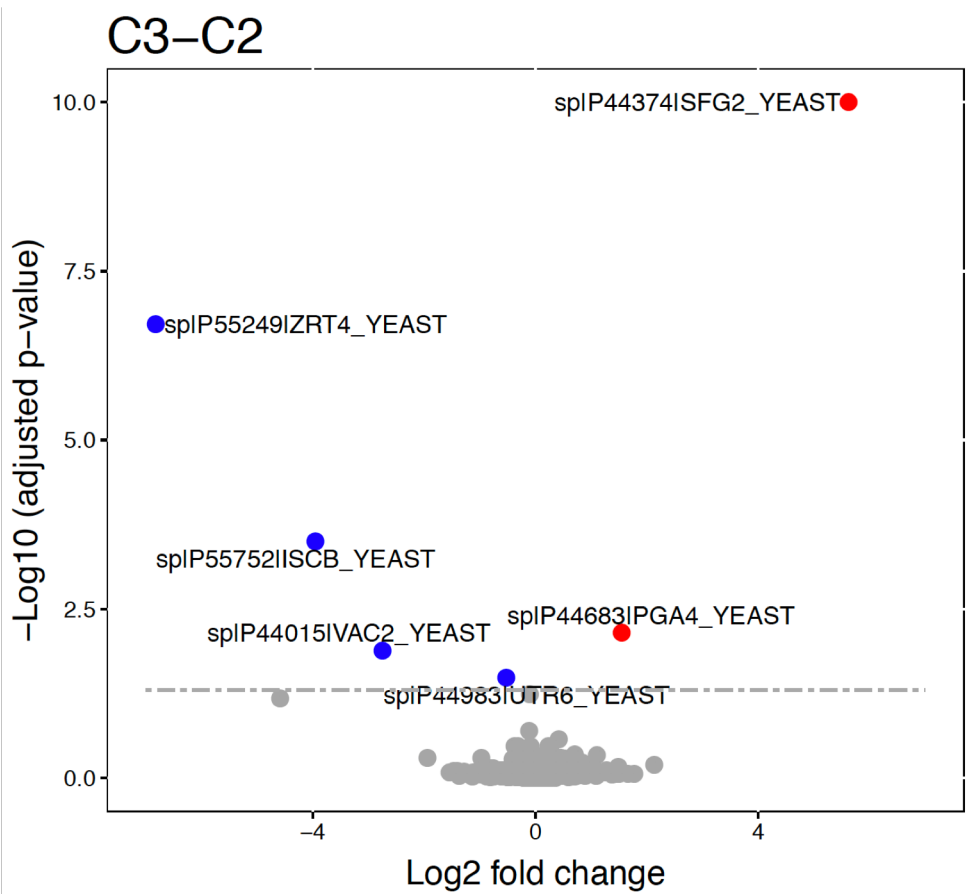
N of significant proteins	Case1-correct	Case2-incorrect
Background protein	1	4
Spike protein	28 out of 30	26 out of 30



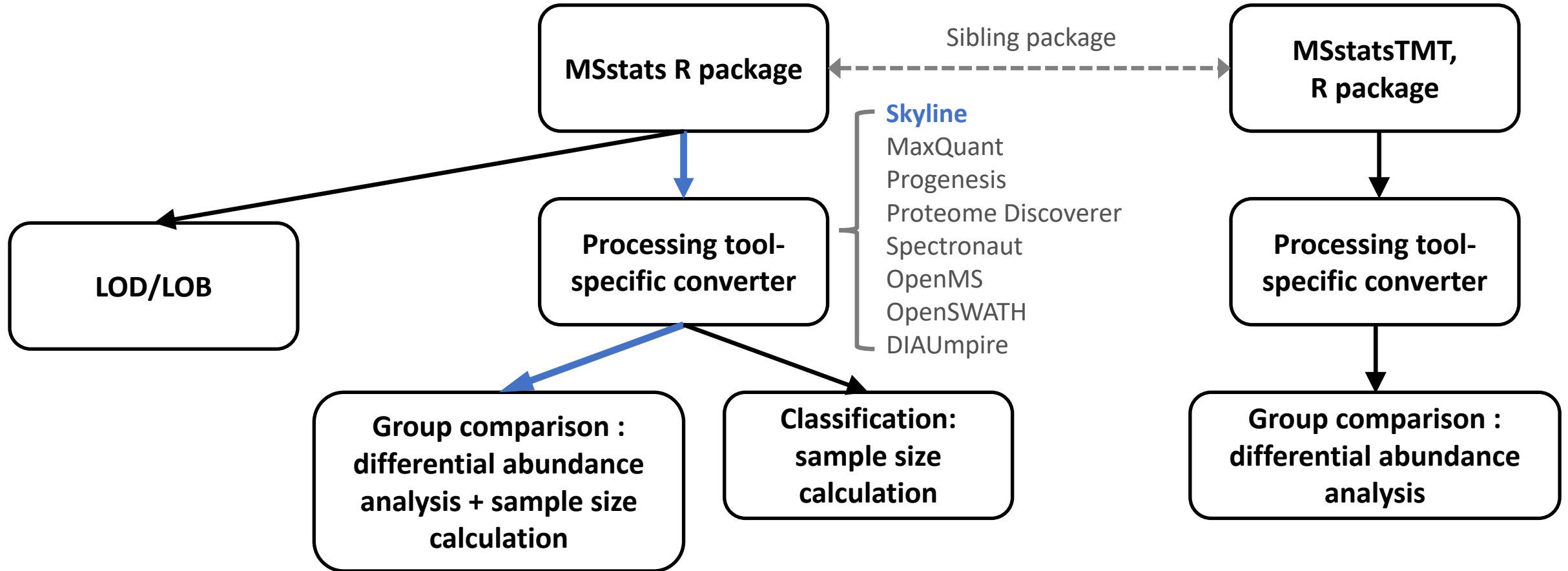


# 3. Different parameter settings have different results

N of significant proteins	Case1- iProphet cutoff=0.15	Case2- iProphet cutoff=0.05
Background protein	1	5
Spike protein	28 out of 30	28 out of 30



# 4. Other workflow with MSstats R package



\* All materials for each workflow are available in [msstats.org](https://msstats.org)